DOCUMENT RESUME

ED 159 203 95 TM 007 482

AUTHOR Kugle, C. L.
TITLE A Potential Source of Bias in Classroom Observation Systems: Coder Drift.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Mar 78
CONTRACT NIE-C-74-0088
GRANT OB-NIE-G-0116
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978); Pages 28 through 29 of the original document are copyrighted and therefore not available. They are not included in the pagination

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Behavior Rating Scales; Bias; Classroom Observation Techniques; *Codification; *Error Patterns; *Evaluators; Higher Education; *Reliability; Secondary School Teachers; *Teacher Behavior
IDENTIFIERS Flanders System of Interaction Analysis; *Interrater Reliability

ABSTRACT

The possibility that "coder drift" may be a source of bias in classroom observation systems was investigated in two parallel experiments using the Flanders System of Interaction Analysis. Coder drift may be described as the differences in coding which result when observers create their own idiosyncratic systems of classifying teacher behaviors. In this experiment, observers all underwent the same training and then worked in pairs coding teacher behaviors in videotaped junior high school social studies lessons. When within-pair reliabilities were higher than between-pair reliabilities, coder drift was judged to have occurred. Results of analyses on the relationship between coding-related talk between coders within pairs and similarities in their coding, analyses on the differences in coding which resulted when coders worked alone, and discriminant analyses on differences among pairs of coders, are reported. The discussion focuses on the implications of these results for classroom research on teacher effects. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED159203

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Potential Source of Bias in Classroom Observation Systems: Coder Drift

C. L. KUGLE

Research and Development Center for Teacher Education

The University of Texas at Austin

TM007 482

This study was supported in part by the National Institute of Education Contract No. NIE-C-74-0088, Grant No. OB-NIE-G-0116, Research and Development Center for Teacher Education. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that office should be inferred. Paper presented at the annual meeting of the American Educational Research Association held in Toronto, March 27-31, 1978.

Abstract

Coder drift, a potential source of bias in classroom observation systems, exists when systematic differences in coding occur as coders rate the same teacher behaviors. This effect was investigated in a study in which subjects, trained on the Flanders Interaction Analysis System, coded videotapes of teachers. Results of analyses on the relationship between coding-related talk between coders and similarities in their coding, analyses on the differences in coding which result when coders code alone, and discriminant analyses on differences among pairs of coders, are reported. Discussion focuses on the implications of these results for teacher effects research.

A Potential Source of Bias in Classroom Observation Systems: Coder Drift

For many years, a substantial portion of teacher effects research energy has been focused on the nature of the relationship between what a teacher does during the instructional process and how much pupils gain academically as a result of that process. This type of investigation, known as a process-product study, typically attempts to relate measures of pupil achievement to measures of teacher behavior. The teacher behavior data usually consist of ratings by trained observers, working singly or in pairs, on one or more classroom observation systems, which are instruments composed of a standardized set of behavioral categories. Although the process-product approach to the investigation of education seems a worthy avenue to productive conclusions, in general these studies have not indicated any clear-cut relationships between teacher behavior and student achievement upon which a "science of teaching" might be built. Although the reasons for this lack of findings may lie in a variety of areas, there has been a recent focus on the possibility that the classroom observation systems used to quantify teacher behavior in process-product studies may be subject to bias. The intent of the present report is to explore one suggested source of bias in classroom observation systems, that of coder drift.

Coder drift may be described as the differences in coding which result when observers create their own idiosyncratic systems of classifying those teacher behaviors which cannot be readily identified as belonging to a particular category of the classroom observation system being used. Although most observation systems are constructed to contain

mutually exclusive and exhaustive categories, and potential coders are provided with training materials containing numerous examples of which teacher behaviors go in which categories, in reality there are a large number of teacher behaviors which can occur and should be coded, and which do not readily correspond to only one category definition. Considering the situation where coders desire consistency in their ratings, are working in pairs coding the same classrooms, and must make as many as 20 coding decisions in one minute, there is a natural tendency to discuss ways to best categorize ambiguous teacher behaviors. For example, one pair of coders may decide that a particular type of teacher utterance belongs in the category "teacher lectures," while another two coders working together decide the same type of utterance should be classified as "teacher accepts and uses ideas of students."

Preliminary evidence for this type of coder interaction was found in an earlier study on classroom observation systems, (see Kugle, 1977, for a detailed description of this study) which involved six pairs of coders, each utilizing a different classroom observation system, who viewed and coded 36 hour-long videotapes of junior high school teachers. After preliminary practice sessions in coding videotapes, coders were given feedback on their reliabilities, and encouraged to achieve and maintain a high level of coding consistency. During postcoding interviews, the coders indicated that they had had extensive discussions regarding difficult coding events. These discussions involved the identification of key words as signals for the appropriate code to use and the elaboration of lengthy "ground rules" which helped steer coding toward unanimity. For example, one pair of coders decided that the teacher utterances

"wow," "great," and "very good" should be considered supporting, while the teacher utterances "fine," "yes," "right," and "all right," should be considered approving. Further, the coders indicated that these ground rules and key words were felt to be fairly arbitrary divisions in the codes and could have been established in a different manner by another pair of coders using the same classroom observation system. Consideration of this information led to the conclusion that consistent differences in coding would result from these types of decisions, which would in turn produce different pictures of the teacher being observed. In the typical process-product study, however, consistent differences in coding are not detectable, since all observers, or all pairs of observers, do not all rate the same teacher at the same time and thus comparisons of coding frequencies are not meaningful. In order to make meaningful comparisons of consistent differences in coding (labeled coder drift), the following paradigm was developed.

A number of observers undergo the same training procedures on a particular classroom observation system, then work in pairs coding the same teachers during the same lesson (via videotapes). When all pairs code the same situations using the same coding system and differences in interrater reliabilities occur such that between-pair reliabilities are high and among-pair reliabilities are low, this is evidence that the coders are classifying the observed teacher behaviors differently from other pairs but similarly to their partners. In addition, if the coder pair is separated and the individuals code the same tapes in separate rooms, evidence for the coder drift effect may be observed as between-pair reliabilities drop across time. These methods of analyzing coder

drift for the data of the current report are reported elsewhere (Marston, Zimmerer, & Vaughan, 1977); the objectives of the present report will be to present additional sources of support for the hypothesis that coders who work closely together resolve ambiguous categorizations in idiosyncratic ways, and to present a closer look at which particular components of the classroom observation system are susceptible to the coder drift effect. These sources of support will be based on analyses of the data obtained directly from the coding sessions, without transforming the data to reliability coefficients. It was felt that more meaningful and direct interpretation of results could be obtained in this way. In addition, a finer-grained examination of how the coder drift phenomenon occurs is attempted in these analyses which, it is hoped, will allow a clearer picture of where ambiguities in the coding system lie and how they are resolved among coder pairs. The additional sources of evidence for coder drift to be presented in this report are: (1) results of analyses on the relationship between the amount of coding-related talk which occurred between pairs of coders working together and the differences in their coding; (2) results of analyses on differences in coding between pairs of coders who worked together and coders who coded alone; and, (3) results of discriminant analyses on the differences in coding among coder pairs working together. The data for these analyses come from an experiment utilizing the paradigm described above, which was conducted specifically to investigate the coder drift effect.

Method

Coders

Twenty-eight paid student volunteers were recruited mainly through advertisements at the financial aid office of The University of Texas at

Austin. They were paid $4 per hour of training and testing with a $20 bonus for completion of the test series. Two coders had insufficient time periods available to be scheduled so only 26 coders completed the experiment.

Instruments

The classroom observation system used in this experiment was the Flanders System of Interaction Analysis (FSIA) (Flanders, 1970), which contains ten categories of classroom verbal behavior. Coding on this instrument is done at 3-minute intervals, the observer indicating by a tally mark in the appropriate category which type of verbal behavior occurred during the interval. Scoring on the FSIA is in terms of frequencies for the verbal behavior categories which occurred during the lesson. A copy of this instrument is included as Appendix A.

Training

Training in the use of the FSIA was given in two 3 1/2 hour sessions on consecutive evenings. The training was based on the audio cassette method provided by Multi-Media Associates. Some minor modifications had to be made to accommodate the large group. In particular, it was not possible to give any feedback on the coders' reliabilities since they did not have partners at this time. Because of this, considerable emphasis was placed on the discussion of the correct coding for the training tapes. At the end of each training session the coders made an actual coding of a tape from the same series that they would be using for the test part of the experiment. Reliabilities between all possible pairs of coders were computed for these two tapes and averaged. These reliabilities were then used to assign coders to pairs so that all pairs were at approximately the same level of coding performance at

the beginning of the experimental sessions.

Materials

The coders viewed the test tapes in one of two office-sized rooms equipped with a 48 cm (19 inch) black and white television monitor, a tone signal, a microphone, a button to signal the experimenter, and a work table. Two conditions were utilized, labeled "separate" and "together." Under the separate condition, members of a pair coded the same tape at the same time in different rooms, while under the together condition either a single pair or two pairs of coders worked at the same time in the same room. The coders were allowed to adjust the television monitor to their own taste and could signal the experimenter with the button if any problem developed. The coders made their codes by checking boxes under the appropriate behavior category on standardized coding forms.

The experimenter presented the videotapes using a Sony VP-2000 U-matic 19 mm (3/4 inch) cassette videoplayer. Tone signals recorded directly onto the video cassettes marked the end of 10-minute time intervals and indicated that coders were to start on a new coding form. An audiotape recording was made of everything the coders said from the time they entered the coding room until they left in order to obtain a quantifiable record of coding-related talk between members of a pair.

The tapes coded were from a videotape library supplied by the Stanford R&D Center; they consisted of ten hour-long sessions of teachers instructing junior high school students in social studies. Two videotapes from this library, similar in style to the first videotapes, were used during training for practice in coding and to allow an assessment of reliabilities at the end of training.

Method and Design

The coders were assigned to rate the experimental tapes in one of two groups. In the first group coders worked with their partners for the first five tapes and then worked in separate rooms for the last five tapes. In the second group, partners came in at the same time but worked in a separate room for the first five tapes; these coders then worked together for the last five tapes. The coders were encouraged by the experimenters to be as accurate as possible but were never told that they should maintain a high reliability with their partner. If the coders asked the experimenters any questions about the FSIA they were told to refer to the materials they had received during the training. Any questions about the experiment were deferred until the debriefing session, which was conducted after tape 10 had been coded.

The coder pairs were formed using a combination of the reliabilities from the two training videotapes, and schedules of available times provided by the coders. This was done by first calculating a Scott's PI (Scott, 1955) coefficient between every possible pair of coders for each of the two training tapes. These reliabilities were then transformed using the arcsin method for proportions and averaged. Pairs that had less than 20 sessions in common on their time schedules were dropped from consideration. Finally, a trial-and-error procedure was used to find a set of pairs that had training reliabilities close to the overall average and that showed the least variance in reliability between pairs. Seven pairs were randomly assigned to each of the groups and scheduled so that with the exception of a few cases they would never see more than one tape a day. One pair had to be dropped after the first six tapes in

the together-apart group, leaving a total of 26 subjects making 13 pairs. All subjects saw the videotapes in the same sequence (see Zimmerer & Kugle, 1978, for a detailed presentation of this study).

Results

Each of the research questions under consideration in this report will be treated separately. As noted previously, the data collected during these experiments were in the form of frequencies of teacher behaviors which occurred in each of the ten categories on the FSIA during each videotape session. The raw frequencies were prepared for analysis in a manner appropriate to the research question under consideration and will be described separately for each section. It should be noted that although there are ten categories on the FSIA, categories 1 (teacher accepts feeling of student) and 7 (teacher criticizes or justifies authority) occur with such low frequency throughout the videotapes that some analyses on these categories could not be conducted; this will be noted where appropriate.

Is the Amount of Coding-related Talk Between Coders Related to Their Similarity in Coding?

If coders who work in the same room discuss the categorization of difficult-to-code teacher behaviors, then it would be expected that their interpretation of the teacher behaviors would become more similar. Thus, one would expect that the amount of coding-related talk which occurred between members of a pair would be negatively correlated with the differences in coding between them. In other words, the more coders talk about categorization, the smaller the differences in their coding.

In order to investigate this possibility, the data from the FSIA were prepared as follows. A total frequency was obtained across all categories for each coder, for each videotape. These individual frequencies for each category were then divided by the total frequency for the videotape, thus converting the category frequencies to proportions of category codes per videotape session. This eliminated any differences in coding due to different rates of coding by the subjects. Finally, the absolute differences in proportions between members of a pair for each category, for each videotape, were obtained.

As mentioned previously, an audio recording of comments made by coders during the viewing of the videotapes was made. Transcripts of these audiotapes were made and scored for the amount of coding-related talk which occurred between coders. These frequencies were then correlated with the absolute differences in category proportions between coders, for those tapes in which coders worked together.

The results of these analyses are presented in Table 1. Significant correlations were obtained for tapes 4 and 5, Category 3 (teacher accepts or uses ideas of students). Although this number of significant correlations is fewer than would be expected by chance, it is clear from examination of Table 1 that the directionality of the hypothesis (i.e., negative correlations), was supported. A sign test performed to determine if the number of negative correlations was greater than expected by chance confirmed this expectation, ($p < .01$).

Do Coders Who Work Together Demonstrate Smaller Differences in Their Categorizations of Teacher Behavior Than Coders Who Work Alone?

In order to answer this question, analyses of variance were performed on the absolute differences in frequency for each category between pairs

of coders. In these analyses, the dependent variables were the frequencies on the 10 categories of the FSIA. The independent variables were those of Group (Together-Apart and Apart-Together), Pair (pairs 1-6 for Group 1, and pairs 7-14 for Group 2), order (either Together-Apart for the coding, or Apart-Together) and tape (10 sessions).

The tests of interest in these analyses are those of the main effect for group and the interaction of group by order. As mentioned above, order refers to the condition of coding utilized in the design of the experiment, that is: half of the pairs of coders (Group 1) worked together for the first five tapes and apart for the last five tapes while the other half of the pairs (Group 2) had the reverse order of coding condition. If coder drift occurred among the subjects, then the differences in coding should show an increase for Group 1 (Together-Apart) from the first to the second condition and a decrease for Group 2 (Apart-Together) from the first to the second condition.

The means for the groups, the F- values, and the probabilities from their analyses are presented in Table 2. Significant results (α = .10 and .05) are indicated by an asterisk. As can be seen from the table, the main effect for group was significant for Category 5 (teacher lectures). The Groups x Order effect was significant for Category 6 (teacher gives directions); Category 8 (student talk-response); and, Category 9 (student talk-initiation).

For those categories that showed significant Group x Order interactions, all of the mean values were in the expected direction: the differences in coding increased when coders worked separately and decreased when they worked together. For Category 5, which showed a

significant group difference, the overall mean differences in coding for Group 2 were larger than for Group 1. Although differences in coding appear to diminish as coders work together, as seen from the interaction means, it appears that working separately to begin with produces differences that do not readily dissipate under the together condition, thus resulting in overall larger differences in coding for Group 2 as compared to Group 1.

Support for these results has been received in the analyses of data from a second experiment (Experiment II) conducted to examine the coder drift effect. This experiment was essentially a replication of the study described in this report with the following modifications. The videotapes coded were of fifth grade teachers in relatively traditional style classroom settings and were approximately a half-hour each in duration, as opposed to the hour-long videotapes of the relatively open-style junior high school classrooms of the first experiment. In addition, the coder pairs in Experiment II saw the tapes in a randomized sequence, rather than in the same sequence which was the procedure for Experiment I.

The independent and dependent variables for these analyses were the same as those for the analyses of the data from Experiment I. Results of the analyses of variance on the data from Experiment II are presented in Table 3. Significant group differences were obtained for Category 4 (teacher asks questions), Category 5 (teacher lectures), and Category 9 (student talk-initiation). Again, the means for Group 1 were lower than for Group 2, indicating smaller differences in coding across all tapes for the together-apart condition than for the apart-together condition.

Significant Group x Order interactions were obtained for Category 5 (teacher lectures), Category 6 (teacher gives direction), Category 8

(student talk-response), and Category 10 (silence or confusion). In all cases, the means for the groups increased during the apart condition (indicating greater differences in coding) and decreased during the together condition.

The results of the analyses for Experiment II provide support for those of Experiment I--particularly with regard to Categories 5, 6, 8, and 9. The fact that these categories are replicated with significant results across the two experiments provides evidence for their particular susceptibility to the conditions of coding.

Do Pairs Within a Group Code Differently from Other Pairs in the Group?

Assuming that there are a variety of ground rules and rationalizations which can be applied to ambiguous coding situations, it seems a reasonable speculation that coders can "drift," as it were, in several ways. That is, members of a pair may drift apart from one another when coding separately, and pairs within a group may drift apart from other pairs within that group. The analyses of variance indicated that members of a pair do tend to have more similar proportions of codes when they code together and less similar proportions when they code alone. It would be of interest to determine whether the pairs within the group coding together can be separated from one another on the basis of their proportions of codes, which would indicate that separate coding "sets" can arise due to different but consistent resolutions of ambiguous behaviors.

One method of analysis which allows examination of these kinds of groupings is that of discriminant analysis. The function of a discriminant analysis is to weight and linearly combine characteristics of various groups so that the groups are made to be as statistically distinct as

possible. In this case, only the data from the coding-together condition were used, and two separate analyses were run. Thus, the groups in the first discriminant analysis were those six pairs who coded together for the first five tapes (these are the pairs from the Group 1, together-apart condition), and the discriminating variables were the proportions of codes for each category across the five tapes. As in the first discriminant analysis, the second discriminant analysis utilized the seven pairs who coded together for the last five tapes (Group 2, the apart-together condition) as the groups and the proportions of codes across the five tapes as the discriminating variables.

Results of the discriminant analyses are presented in Table 4. Overall multivariate F-tests of the differences among the group centroids, with associative probability values, are given, as well as the chi-square tests which determine how many discriminant functions are adequate for distinguishing among the groups. Overall significant group differences were obtained for Category 8 (student talk-response); Category 9 (student talk-initiation); and Category 10 (silence or confusion) for Group 1, and Category 6 (teacher gives directions: Category 7 (teacher criticizes pupil); and Category 8, for Group 2. Again it may be seen that the categories which were associated with significant results in the analyses of variance (Categories 6, 8, and 9) are related to significant group differences in these analyses as well.

Further evidence about the group differences can be obtained from examination of the group centroids, which are the mean discriminant scores for each group on the respective discriminant functions. The centroids represent the most typical location of a member from that group in the

discriminant function space. Although these locations might be more easily visualized in graphical form, there are more than two discriminant functions derived for most of the categories with significant group differences, making graphical representations difficult. However, Category 8 for Group 2, which has two significant discriminant functions, has been plotted to provide a clearer presentation of the separation of pairs within the group, and is presented in Figure 1. The location of the group centroid is indicated by an asterisk, and the discriminant scores for the individual members of the pair are indicated by the number of the pair. It can be seen from this figure that the coder pairs are located relatively close together, except for pairs 4 and 6, who differ both from the other pairs and from one another.

In addition to these group separations, it is possible to examine the contributions of the discriminating variables to the discriminant functions, presented as correlation coefficients in Table 5. Thus, on Category 7--Group 2, for example, the first function primarily represents differences in coding on the second videotape which the pairs coded under the together condition, while the second function is reflective of the coding proportions on the fifth tape coded during the together condition. The contributions of the discriminating variables to the discriminant functions are readily apparent in most cases and appear to indicate that differences in the behaviors on the tapes contribute to the separation of the groups.

This interpretation is supported by the results of univariate F-tests performed on the discriminating variables. These are presented in Table 6, with their associative probability values. For example, from the table it can be seen that pair differences for Category 8--Group 1, are significant

across all five of the discriminating variables. This indicates that there are significant differences in the proportions of codes for Category 8 for all of the videotapes which the coders viewed during the together condition. Whether the differences among the pairs are significant for all possible comparisons of pairs, or whether the significant overall differences is attributable to one or two pairs who are classifying the behaviors differently remains to be determined. Answering the question of where these differences exist must await further analyses, and should provide additional information as to how many coding "sets" can be generated from one classroom observation system. For the present, the results of the discriminant analyses indicate that pairs do differ in their coding of the videotaped lessons. Thus, while members of a pair may show consistency in coding with one another, it appears that differences exist across pairs in the interpretation of what behaviors are occurring during the instructional process.

Discussion

The purposes of the present report were to provide additional support for the coder drift effect (reported in Marston, et al., 1978), as well as additional information regarding how this effect occurs, and in what ways it may be manifested.

The first section of this report dealt with how coders through discussion of ambiguous coding events become more similar in their classification of the behaviors coded. Given that all possible teacher actions cannot be discussed during training, it is intuitively obvious that behaviors which do occur and are not readily classifiable would be a topic of coder conversation. It also seems apparent that such

conversations would lead to an acceptance of common ground rules for coding. Although the number of significant correlations obtained between the amount of coder talk and the differences between coders in coding was not impressive, the fact that a significant number of them were in the right direction was encouraging. Since the number of coders in this analysis was no more than 14, it is felt that a larger sample would provide a greater likelihood of obtaining significant correlations, and replication of this analysis will be conducted for the data from Experiment II. It is anticipated that the results of the replication will support the relationship between coder talk and coding similarities.

Results from the second section of this report provided support for the coder drift effect, indicating that working together reduces differences in coding, while working apart increases them. It also appears that the order of condition of coding may affect the differences in coding between members of a pair. That is, coding with a partner initially seems to provide overall lower differences in coding within a pair than coding alone for a number of tapes before coding with a partner. It may be that coders who code alone initially resolve ambiguous situations in their own way and are less willing to be influenced by a partner later in coding. This is in contrast to coders who are faced with an ambiguous situation early in coding with another coder present with whom they can discuss coding difficulties.

Results from the third section of this report indicated that pairs may drift apart from one another, suggesting that, for some categories at least, different classificatory rules are in operation depending on the pair of coders doing the classification.

Overall, the results of these analyses indicate that some categories on the FSIA seem to be more susceptible to coder interpretation than others. In particular, Category 6 (teacher gives directions) and the two pupil categories--Category 8 and Category 9--are regularly associated with significant group differences throughout this report. Given that only eight categories were consistently examined (due to the low frequency of Categories 1 and 7) it is fairly impressive that three categories out of the eight examined appear with regularity to be associated with differences in coding. This has important implications for research in the teacher effects field. If some categories of classroom behavior are susceptible to fluctuations in coding due to the interpretations of the coders, this introduces a considerable amount of error in the data collected. Thus, it is no wonder that so few consistent process-product relationships are obtained in teacher effects research.

Analyses of the data from Experiment II are currently underway and should provide further insight into the ways in which differences in coding occur; ultimately it is hoped that such analyses will identify the sources of the coder drift phenomenon and provide a means for strengthening classroom observation systems so that our analyses of teacher behavior will be more accurate and thus more powerful.

Reference Note

[1]The expert trainer and the training materials were provided by EPIC Systems, 4901 East Fifth Street, Tucson, Arizona 85132.

References

Flanders, N.A. Analyzing Teaching Behavior. Reading, Massachusetts: Addison-Wesley, 1971.

Kugle, C.L. Data collection procedures for the Evaluation of Teaching Program, Phase II. (Tech. rep.) Austin, Texas: The University of Texas Research and Development Center for Teacher Education, 1977.

Marston, P.T., Zimmerer, L.K., & Vaughan, W.C. Coder drift: a reliability problem for teacher observations. Paper presented at the American Educational Research Association Annual Meeting, Toronto, Canada, 1978.

Scott, W.A. Reliability of content analysis: the case of nominal scale coding. The Public Opinion Quarterly, 1955, 19, 321-325.

Zimmerer, L., & Kugle, C.L. Data collection procedures for the Evaluation of Teaching Program: Coder Drift Studies. (Tech. rep.) Austin, Texas: The University of Texas Research and Development Center for Teacher Education, 1978.

Table 1

Correlations Among Coding-related Talk and Differences in Coding Between Pairs of Coders, for Those Sessions in Which Coders Worked Together.

	TAPES				
Category	1	2	3	4	5
1	.08	-.31	.35	-.04	.01
2	-.19	-.18	-.17	-.27	-.17
3	.04	-.28	-.25	-.50*	-.51*
4	.19	-.09	.01	-.15	-.12
5	.10	.08	.02	-.02	-.15
6	-.02	-.10	.19	-.03	-.30
7	.02	-.12	-.32	-.30	-.10
8	.07	-.09	-.02	-.24	.20
9	-.41	-.23	-.15	-.14	-.22
10	-.27	.21	.03	-.34	-.25

*$p < .05$, df = 11, one-tailed test

Results of Sign Test:

For N = 50, a value less than or equal to 15 for the number of positive correlations obtained is significant at the .01 level. For the data in Table 1, the number of positive correlations obtained equals 15.

Table 2

Results of Analyses of Variance on Absolute Differences in Frequency of Coding for Each Category Between Pairs of Coders, For Experiment I.

Category		Tapes 1-5	Tapes 6-10	Group $\bar{X}$
1.	GP1	.2933	.0333	.1633
	GP2	.4743	.1257	.3000
	G	F"(2,13)=2.45, p=.125		.2317
	GO	F"(17,14)=1.28, p=.323		
2.	GP1	1.113	1.393	1.253
	GP2	1.903	1.583	1.743
	G	F"(1,16)=1.61, p=.223		1.498
	GO	F"(2,17)=.98, p=.395		
3.	GP1	8.513	7.173	7.843
	GP2	11.526	6.949	9.237
	G	F"(2,14)=.61, p=.555		8.540
	GO	F"(1,18)=1.96, p=.178		
4.	GP1	5.913	5.707	5.810
	GP2	12.617	4.914	8.766
	G	F"(1,13)=.96, p=.344		7.288
	GO	F"(1,15)=3.03, p=.102		
5.	GP1	10.09	13.88	11.98
	GP2	18.12	19.29	18.70
	G	F"(1,19)=3.22, p=.089*		15.34
	GO	F"(33,18)=.50, p=.960		
6.	GP1	3.427	8.013	5.720
	GP2	5.880	4.989	5.434
	G	F"(19,18)=.22, p=.999		5.577
	GO	F"(1,19)=6.13, p=.023**		
7.	GP1	.1667	.5200	.3433
	GP2	.2000	.3663	.2831
	G	F"(6,19)=.51, p=.791		.3132
	GO	F"(3,15)=1.21, p=.341		
8.	GP1	9.660	21.467	15.563
	GP2	31.720	20.406	26.063
	G	F"(1,19)=2.88, p=.106		20.813
	GO	F"(1,14)=5.11, p=.040**		
9.	GP1	9.947	29.413	19.680
	GP2	36.543	14.777	25.660
	G	F"(2,19)=1.21, p=.319		22.670
	GO	F"(1,17)=15.17, p=.001**		
10.	GP1	6.160	10.333	8.247
	GP2	13.320	7.920	10.620
	G	F"(2,18)=.63, p=.544		9.433
	GO	F"(1,18)=2.15, p=.160		

GP1 = Group 1 (Together-Apart)

GP2 = Group 2 (Apart-Together)

G = Main Effect for Groups

GO = Interaction Effect-Group x Order

Table 3

Results of Analyses of Variance on Absolute Differences in Frequency of Coding for Each Category Between Pairs of Coders, For Experiment II.

Category		Tapes 1-5	Tapes 6-10	Group $\bar{X}$
1.	GP1	.0444	.1222	.0833
	GP2	.0857	.0667	.0762
	G	F"(82,15)=.93, p=.613		.0796
	GO	F"(2,19)=1.42, p=.267		
2.	GP1	.856	1.044	.950
	GP2	1.248	2.324	1.786
	G	F"(1,12)=.90, p=.362		1.368
	GO	F"(2,15)=.86, p=.443		
3.	GP1	6.833	10.189	8.511
	GP2	11.857	14.952	13.405
	G	F"(1,13)=2.72, p=.123		10.958
	GO	F"(89,16)=.40, p=.997		
4.	GP1	7.056	7.556	7.306
	GP2	12.267	10.771	11.519
	G	F"(1,18)=3.39, p=.082*		9.412
	GO	F"(5,11)=1.20, p=.371		
5.	GP1	10.02	12.16	11.09
	GP2	19.51	12.73	16.12
	G	F"(1,17)=4.64, p=.046**		13.61
	GO	F"(1,16)=3.54, p=.078*		
6.	GP1	7.800	9.056	8.428
	GP2	9.190	5.581	7.386
	G	F"(2,16)=.86, p=.440		7.907
	GO	F"(1,19)=8.25, p=.010**		
7.	GP1	.3111	.1444	.2278
	GP2	.3048	.2762	.2905
	G	F"(8,18)=.67, p=.715		.2591
	GO	F"(7,16)=.85, p=.561		
8.	GP1	10.51	18.59	14.05
	GP2	21.66	13.34	17.50
	G	F"(2,19)=.75, p=.484		16.02
	GO	F"(1,19)=4.67, p=.044**		
9.	GP1	3.222	4.789	4.006
	GP2	12.229	8.981	10.605
	G	F"(1,18)=6.38, p=.021**		7.305
	GO	F"(3,17)=1.19, p=.345		
10.	GP1	4.100	4.944	4.522
	GP2	7.924	3.819	5.871
	G	F"(2,19)=1.11, p=.350		5.197
	GO	F"(1,19)=4.30, p=.052**		

GP1 = Group 1 (Together-Apart)

GP2 = Group 2 (Apart-Together)

G = Main Effect for Groups

GO = Interaction Effect-Group x Order

Table 4

Discriminant Analyses: Results of F-Tests for Overall Significant Group Differences, and Chi-Square Tests for Significant Discriminant Functions (Roots)

Category	Group 1, Pairs 1-6	Group 2, Pairs 7-14
1	$F(25,9) = 0.849$, $p = .6496$ No significant functions	See Footnote.
2	$F(25,9) = 1.497$, $p = .2706$ Root 1: $\chi^2(9) = 19.682$, $p = .0207$ Root 2: $\chi^2(7) = 14.783$, $p = .0394$	$F(30,14) = 2.176$, $p = .0627$ Root 1: $\chi^2(10) = 28.036$, $p = .0023$ Root 2: $\chi^2(8) = 16.571$, $p = .0356$
3	$F(25,9) = 2.617$, $p = .0683$ Root 1: $\chi^2(9) = 27.144$, $p = .0018$	$F(30,14) = 0.737$, $p = .7666$ No significant functions
4	$F(25,9) = 2.801$, $p = .0560$ Root 1: $\chi^2(9) = 25.965$, $p = .0026$ Root 2: $\chi^2(7) = 16.010$, $p = .0256$	$F(30,14) = 0.963$, $p = .5548$ Root 1: $\chi^2(10) = 19.762$, $p = .0325$
5	$F(25,9) = 0.814$, $p = .6780$ No significant functions	$F(30,14) = 1.426$, $p = .2436$ Root 1: $\chi^2(10) = 22.988$, $p = .0115$
6	$F(25,9) = 1.699$, $p = .2063$ Root 1: $\chi^2(9) = 24.785$, $p = .0038$	$F(30,14) = 5.331$, $p = .0013$* Root 1: $\chi^2(10) = 39.926$, $p = .0001$ Root 2: $\chi^2(8) = 19.851$, $p = .0116$
7	See Footnote.	$F(30,14) = 3.274$, $p = .0114$* Root 1: $\chi^2(10) = 30.182$, $p = .0012$ Root 2: $\chi^2(8) = 22.973$, $p = .0040$
8	$F(25,9) = 11.729$, $p = .0007$* Root 1: $\chi^2(9) = 31.952$, $p = .0004$ Root 2: $\chi^2(7) = 25.797$, $p = .0009$ Root 3: $\chi^2(5) = 19.506$, $p = .0020$	$F(30,14) = 2.766$, $p = .0238$* Root 1: $\chi^2(10) = 34.463$, $p = .0003$ Root 2: $\chi^2(8) = 16.619$, $p = .0350$
9	$F(25,9) = 16.376$, $p = .0003$* Root 1: $\chi^2(9) = 47.413$, $p = .0000$ Root 2: $\chi^2(7) = 24.211$, $p = .0015$ Root 3: $\chi^2(5) = 14.350$, $p = .0140$	$F(30,14) = 2.012$, $p = .0836$ Root 1: $\chi^2(10) = 25.772$, $p = .0047$ Root 2: $\chi^2(8) = 16.823$, $p = .0327$
10	$F(25,9) = 5.303$, $p = .0074$* Root 1: $\chi^2(9) = 32.495$, $p = .0004$ Root 2: $\chi^2(7) = 19.335$, $p = .0078$ Root 3: $\chi^2(5) = 11.797$, $p = .0379$	$F(30,14) = 1.852$, $p = .1114$ Root 1: $\chi^2(10) = 23.590$, $p = .0095$ Root 2: $\chi^2(8) = 16.398$, $p = .0377$

Note: Due to the low frequency of occurrence of this category, there was not enough variance in the scores to conduct the analyses

Table 5

Correlations of Discriminating Variables with the Significant Discriminant Functions for those Categories with Significant Group Differences

Category	Variable	Group 1 Significant Functions 1	2	3	Group 2 Significant Functions 1	2	3
6	1	F-Test not significant			.46	.38	
	2				.38	.60	
	3				.82	-.24	
	4				.75	.62	
	5				.89	.44	
7	1	Analyses not conducted			.39	-.18	
	2				.73	-.40	
	3				-.03	.12	
	4				-.06	.39	
	5				.39	.80	
8	1	.36	-.23	.75	.55	.46	
	2	.99	-.03	.09	.41	-.17	
	3	.62	-.42	.55	.69	.05	
	4	.62	.37	.27	.77	.56	
	5	.57	.48	.64	.95	.04	
9	1	.29	.69	-.05	F-Test not significant		
	2	.98	-.17	.01			
	3	.45	.54	.71			
	4	-.69	.55	.42			
	5	.70	.49	-.50			
10	1	.26	-.74	-.05	F-Test not significant		
	2	.14	-.43	.21			
	3	.67	-.03	.49			
	4	.80	.55	.14			
	5	.06	.17	.69			

Table 6

Results of Univariate F-Tests on Discriminating Variables

for Categories with Significant Group Differences

Category	Variable	Group 1 F-Ratio	Group 1 P	Group 2 F-Ratio	Group 2 P
6	1	Overall F-Test not significant		1.7662	.2363
	2			5.6724	.0197*
	3			5.0393	.0265*
	4			17.5478	.0012*
	5			42.8098	.0002*
7	1	Analyses not conducted		1.0000	.4920
	2			4.3676	.0377*
	3			.3496	.8891
	4			.8001	.5997
	5			6.6303	.0133*
8	1	5.0190	.0380*	3.6005	.0593
	2	152.4293	.0001*	1.5968	.2757
	3	15.7715	.0030*	2.3700	.1419
	4	5.1234	.0363*	9.9127	.0047*
	5	23.2004	.0014*	13.3575	.0023*
9	1	4.4845	.0483*	Overall F-Test not significant	
	2	132.7067	.0001*		
	3	17.6079	.0024*		
	4	20.0072	.0019*		
	5	24.8032	.0012*		
10	1	2.7318	.1268	Overall F-Test not significant	
	2	0.7141	.6364		
	3	2.4549	.1524		
	4	18.8100	.0021*		
	5	1.7438	.2578		

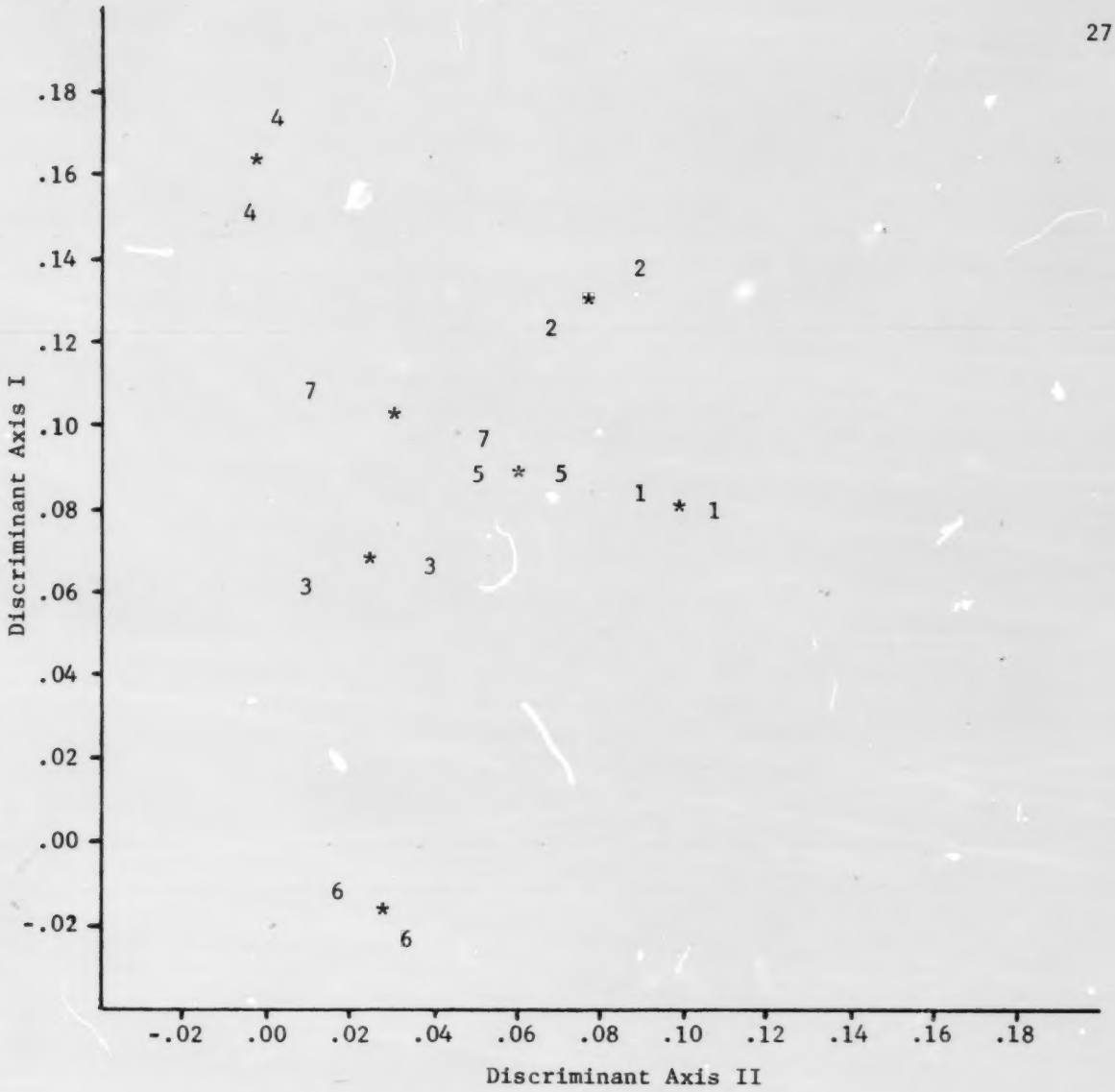

Figure 1. Group centroids(*) and subject points for Category 8, Group 2.

Appendix A, What is Flanders' Interaction Analysis System?, has been removed because of copyright restrictions.